# VOICE BASED SEARCH ENGINE FOR VISUALLY IMPAIRMENT PEOPLES

**Dr. S. Sridharan[1]\*, D. Abinaya[2], M. Jovitha[2]** and **K. Kayathri[2]**

[1] Asst. Prof., University College of Engineering, Thirukkuvalai, Nagapattinam (Dt), Tamil Nadu, India.

[2] Student, University College of Engineering, Thirukkuvalai, Nagapattinam (Dt), Tamil Nadu, India.

## ARTICLE INFO

## ABSTRACT

World Wide Web (WWW) is unexpectedly emerging because the accepted records supply for our society. The WWW is normally reachable the usage of an internet-browsing package from a networked pc. The layout of facts on the net is visually orientated. The reliance on visible presentation locations excessive cognitive demands on a person to function this sort of system. The interaction might also sometimes require the whole attention of a consumer. The design of information presentation at the web is predominately visible-oriented. This presentation technique requires most, if no longer all, of the person's attention and imposes significant cognitive load on a user. This technique isn't always sensible, in particular for the visually impaired persons. The awareness of this challenge is to develop a prototype which supports net browsing the use of a speech-based interface, e.g. A telephone, and to degree its effectiveness. The command input and the delivery of web contents are totally in voice. Audio icons are constructed into the prototype so that users will have higher knowledge of the original shape/purpose of a web page. Navigation and manage commands are available to decorate the net browsing enjoy. The effectiveness of this prototype is evaluated in a consumer take a look at involving both generally sighted and visually impaired humans. Voice browsers allow human beings to get right of entry to the Web the usage of speech synthesis, pre-recorded audio, and speech reputation. This may be supplemented via keypads and small presentations. Voice may also be supplied as an accessory to standard computing device browsers with high resolution graphical presentations, presenting an on hand alternative to the use of the keyboard or screen, as an instance in cars in which palms/eyes unfastened operation is crucial. Voice interplay can get away the bodily obstacles on keypads and shows as cell devices turn out to be ever smaller. The browser will have an integrated textual content extraction engine that inspects the content of the page to construct a structured illustration. The inner nodes of the structure constitute diverse tiers of abstraction of the content. This enables in easy and bendy navigation of the page so that it will hastily home into gadgets of interest.

# INTRODUCTION

Machine studying is software of artificial intelligence (AI) that gives systems the ability to robotically have a look at and beautify from enjoy without being explicitly programmed. Machine studying specializes in the development of pc programs that could get right of entry to records and use it observe for them. The gadget of studying begins off evolved with observations or statistics, collectively with examples, direct revel in, or education, on the way to search for styles in information and make higher choices inside the future based totally on the examples that we offer. The number one purpose is to allow the computer systems analyze routinely without human intervention or help and adjust moves hence. Machine getting to know algorithms are frequently classified as supervised or unsupervised. Supervised algorithms require a records scientist or data analyst with machine learning abilities to offer each enter and desired output, further to furnishing remarks approximately the accuracy of predictions all through set of regulations education. Data scientists determine which variables, or functions, the version want to analyze and use to expand predictions. Once schooling is entire, the set of rules will practice what was determined out to new records. Unsupervised algorithms do now not need to look at with preferred very last consequences records. Instead, they use an iterative method called deep studying to review statistics and arrive at conclusions. Unsupervised getting to know algorithms also known as neural networks  are used for extra complicated processing obligations than supervised gaining knowledge of systems, such as image recognition, speech-to-text and natural language generation. These neural networks paintings through using combing through millions of examples of training information and automatically identifying regularly subtle correlations amongst many variables. Once skilled, the set of rules can use its financial group of institutions to interpret new records. These algorithms have best emerge as feasible in the age of large records, as they require huge amounts of education data. Machine gaining knowledge of algorithms are regularly classified as supervised or unsupervised.

Supervised device learning algorithms can observe what has been observed in the past to new data the use of labeled examples to count on destiny activities. Starting from the evaluation of a diagnosed training dataset, the mastering set of rules produces an inferred function to make predictions about the output values. The device is able to provide goals for any new enter after enough education. The reading set of guidelines can also compare its output with the precise, supposed output and discover mistakes if you need to adjust the model as a result. In comparison, unsupervised gadget studying algorithms are used while the records used to train is neither categorized nor categorized. Unsupervised mastering studies how systems can infer a characteristic to describe a hidden structure from unlabeled facts. The device doesn't discern out the right output, but it explores the records and might draw inferences from datasets to give an explanation for hidden structures from unlabeled records. Semi-supervised device getting to know algorithms fall somewhere in among supervised and unsupervised learning, for the motive that they use each classified and unlabeled records for training – commonly a small amount of classified records and a large quantity of unlabeled facts. The systems that use this method are able to substantially improve gaining knowledge of accuracy. Usually, semi-supervised getting to know is selected when the obtained classified statistics calls for professional and applicable assets with the intention to teach it / examine from it. Otherwise, acquiring unlabeled

information usually doesn't require extra belongings. Reinforcement system mastering algorithms is a mastering approach that interacts with its surroundings via generating actions and discovers errors or rewards. Trial and mistakes search and no longer on time praise are the most applicable traits of reinforcement mastering. This technique permits machines and software program software entrepreneurs to robotically decide the proper behavior internal a specific context on the way to maximize its overall performance. Simple reward remarks are needed for the agent to learn which motion is exceptional; this is referred to as the reinforcement sign. Machine mastering allows analysis of huge quantities of statistics. While it usually presents quicker, greater accurate outcomes as a way to turn out to be privy to worthwhile possibilities or risky dangers, it can additionally require time past law and assets to teach it properly. Combining machine gaining knowledge of with AI and cognitive technology can make it even greater powerful in processing large volumes of records. User interfaces for software program programs can are to be had a selection of codecs, starting from command-line, graphical, net application, and even voice. While the most popular consumer interfaces include graphical and internet-primarily based packages, every so often the need arises for an opportunity interface. Whether because of multi-threaded complexity, concurrent connectivity, or statistics surrounding execution of the issuer, a chat bot based interface may also moreover fit they want. Chat bots usually offer a text-primarily based customer interface, permitting the man or woman to kind instructions and get preserve of text responses. Chat bots are usually state full offerings, remembering previous instructions (and perhaps even communique) on the way to offer functionality. The basic structure of chat bot application can be shown in fig 1.

**Figure – 1:** *Search Engine Optimization*
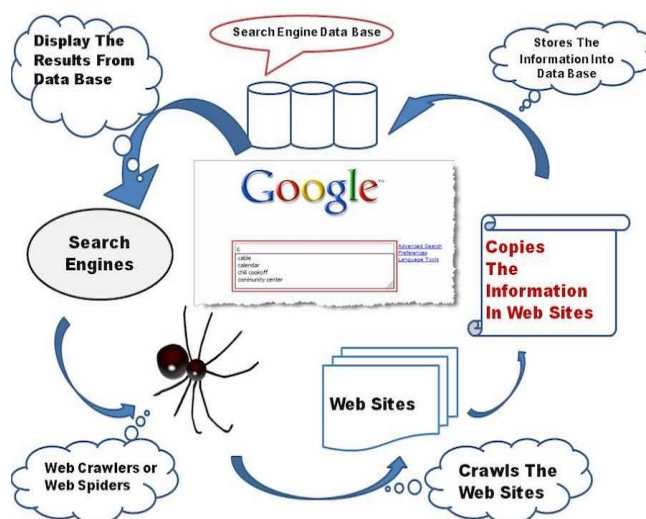


## RELATED WORK

Furui, Sadaoki, *et al.,*[1] Applied the automatic Speech Recognition systems educated on non-prevent speech corpora are used to offer a completely precise mapping between overseas sounds and native lessons. The authors display how the device ABX assessment technique can be used to assess predictions from the resulting quantitative models with empirically attested effects in human cross-linguistic phonetic class notion. In this paper,

we advise to leverage Automatic Speech Recognition (ASR) era to gather absolutely precise mappings between overseas sounds and local training after which use the device ABX assessment task to derive quantitative predictions from those mappings concerning go-linguistic phonetic class perception. More in particular, our approach may be broken down into three steps. First, teach a phoneme recognizer in a "local" language the usage of annotated continuous speech recordings. Second, use the skilled device to derive perceptual representations for test stimuli in a overseas language. In this paper, the ones might be vectors of posterior possibilities over each of the local phonemes. Third, collect predictions for perceptual mistakes by means of taking walks a psychophysical take a look at over those representations for each overseas assessment. Machine ABX discrimination duties might be used for this. To show off the opportunities supplied through way of the method, we study predictions obtained for three empirically-attested consequences in pass-linguistic phonetic magnificence belief. The first results are worldwide results that have a look at to the set of phonetic contrasts in a language as an entire. This representation can take the form of a phonetic class label, a vector of posterior possibilities over possible telephones or a few one of a kind, in all likelihood richer, form of illustration.

Hinton, Geoffrey, *et al.,*[2] Evaluated the wholesome is to use a feed-ahead neural network that takes numerous frames of coefficients as enter and produces posterior possibilities over HMM states as output. Deep neural networks (DNNs) that have many hidden layers and are educated the use of new techniques had been shown to outperform GMMs on a variety of speech popularity benchmarks, from time to time by using a big margin. Artificial neural networks skilled with the useful resource of lower again-propagating errors derivatives have the capability to study a whole lot better models of information that lie on or near a nonlinear manifold. In fact, two a long time in the past, researchers finished a few achievements the use of synthetic neural networks with a unmarried layer of nonlinear hidden units to expect HMM states from domestic home windows of acoustic coefficients. At that point, however, neither the hardware nor the studying algorithms have been good enough for education neural networks with many hidden layers on big portions of facts, and the overall performance benefits of the usage of neural networks with a unmarried hidden layer were now not sufficiently massive too seriously venture GMMs. As a result, the number one sensible contribution of neural networks at that factor becomes to offer more capabilities in tandem or bottleneck systems. GMMs have a number of advantages that lead them to suitable for modeling the opportunity distributions over vectors of enter capabilities which can be related to each country of an HMM. With enough components, they may version opportunity distributions to any required degree of accuracy, and they're pretty easy to inform to data the use of the EM set of rules. A large amount of research has long beyond into finding methods of constraining GMMs to increase their evaluation velocity and to optimize the tradeoff among their flexibility and the quantity of education information required to avoid extreme over fitting.

Real, Esteban, *et al.,*[3] Hired evolutionary algorithms to discover such networks automatically. Despite massive computational requirements, we display that it is now viable to evolve fashions with accuracies in the type of those posted in the final 365 days. In this paper we have tested that (i) neuro-evolution is capable of building massive, accurate networks for two hard and famous photograph magnificence benchmarks; (ii) neuro-

evolution can do this starting from trivial preliminary conditions while searching a completely huge space; (iii) the method, as soon as started out, dreams no experimenter participation; and (iv) the manner yields fully educated fashions. We moreover paid close to attention to result reporting. Namely, we gift the variety in our outcomes similarly to the top cost, we account for researcher stages of freedom, we have a look at the dependence at the meta-parameters, and we divulge the amount of computation vital to attain the primary effects. We are hopeful that our express dialogue of computation price ought to spark greater look at of inexperienced version are looking for and training. Studying version common overall performance normalized with the useful resource of computational investment permits consideration of financial ideas like opportunity price. Our approach builds on previous art work, with a few important variations. We explore huge version-structure seek areas beginning with essential preliminary conditions to avoid priming the tool with information about appeared accurate techniques for the particular dataset available. The encoding isn't like the neuro-evolution strategies referred to above: we use a simplified graph as our DNA, which is transformed to a full neural network graph for training and assessment.

Snoek, Jasper, *et al.,* [4] Evolved gadget gaining knowledge of algorithms with fewer of them. Another, more bendy take in this difficulty is to view the optimization of such parameters as a technique to be automatic. Specifically, we should view such tuning because the optimization of an unknown black-box characteristic and invoke algorithms advanced for such issues. A correct choice is Bayesian optimization, which has been demonstrated to outperform special kingdom of the artwork worldwide optimization algorithms on some of hard optimization benchmark capabilities. For non-prevent skills, Bayesian optimization normally works thru assuming the unknown feature come to be sampled from a Gaussian technique and continues a posterior distribution for this option as observations are made or, in our case, as the consequences of strolling learning set of rules experiments with wonderful hyper-parameters are placed. To select out the hyper-parameters of the following take a look at, you could virtually optimize the anticipated development (EI) over the current top notch result or the Gaussian process top self-notion certain (UCB). The use of machine learning algorithms regularly consists of careful tuning of learning parameters and version hyper parameters. Unfortunately, this tuning is often a "black art" requiring expert revel in, policies of thumb, or sometimes brute force seek. There is therefore terrific attraction for computerized strategies which could optimize the overall performance of any given getting to know set of rules to the problem available. In this work, we take into account this hassle through the framework of Bayesian optimization, wherein a studying set of guideline's generalization performance is modeled as a sample from a Gaussian gadget (GP). We show that wonderful alternatives for the person of the GP, which includes the sort of kernel and the treatment of its hyper-parameters, can play a important position in acquiring a very good optimizer that can gain expert diploma ordinary overall performance. We describe new algorithms that consider the variable fee (length) of getting to know set of rules experiments and which could leverage the presence of more than one cores for parallel experimentation. We show that the ones proposed algorithms enhance on previous automated methods and might reach or surpass human professional-degree optimization for masses algorithms which includes latent Dirichlet allocation, installed SVMs and convolution neural networks.

Shahriari, Bobak, *et al.,*[5] Designed the experiments to benefit insights into bodily and social phenomena, engineers format machines to execute duties greater efficaciously, pharmaceutical researchers layout new drugs to combat sickness, agencies design internet sites to beautify purchaser enjoy and boom marketing income, geologists format exploration strategies to harness herbal assets, environmentalists layout sensor networks to display ecological systems, and builders design software to power pc systems and digital gadgets. All these layout problems are fraught with alternatives, picks which may be often complicated and high dimensional, with interactions that cause them to difficult for human beings to reason about. Big Data programs are commonly associated with structures associated with massive numbers of customers, huge complicated software program structures, and big-scale heterogeneous computing and storage architectures. The creation of such systems involves many allotted layout alternatives. The give up merchandise (e.g., advice systems, clinical assessment system, actual time sport engines, speech recognizers) as a result incorporate many tunable configuration parameters. These parameters are often distinctive and difficult-coded into the software program through numerous developers or agencies. If optimized together, the ones parameters can result in sizable improvements. Bayesian optimization is a powerful device for the joint optimization of format alternatives this is gaining incredible popularity in contemporary years. It ensures extra automation to be able to increase each product quality and human productiveness. This examine paper introduces Bayesian optimization, highlights some of its methodological factors, and showcases a big variety of applications.

**EXISTING METHODOLOGIES**

Voice-enabled interface with addition assist for gesture based totally enter and output strategies are for the "Social Robot Maggie" converting it into an aloud reader. This voice reputation and synthesis can be stricken by number of motives together with the voice pitch, its speed, its quantity etc. It is primarily based on the ETTS (Emotional Text-To-Speech) software program. Robot also expresses its mood thru gesture that is primarily based on geostationary. Speech reputation accuracy may be improved by way of removal of noise. A Bayesian scheme is applied in a wavelet area to separate the speech and noise components in a proposed iterative speech enhancement algorithm.

**Figure – 2:** *Existing Framework*
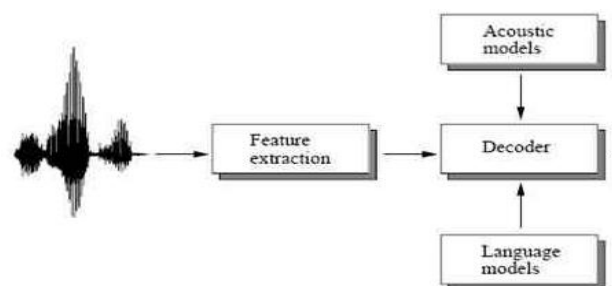
This proposed approach is developed inside the wavelet domain to make the most the chosen functions in the time frequency area illustration. It involves two tiers: a noise estimate degree and a sign separation stage. The Principle Component Analysis (PCA) based totally HMM for the visual modality of audio-visible recordings is used. PCA

(Principle Component Analysis) and PDF (Probabilistic Density Analysis) are modalities records incorporated collectively and acquired a Multi-Stream Hidden Markov Model (MSHMM). MSHMM method is broadly used and very a success in audio visible speech popularity. The current framework only guides educated voices and skilled keywords. The existing framework is shown in fig 2.

## PROPOSED METHODOLOGY

Voice search, additionally referred to as voice-enabled seek, allows the consumer to use a voice command to search the Internet, or a transportable device. Currently, voice search is usually used in (in a slim feel) "directory assistance", or nearby search.  In a broader definition, voice seek include open-domain keyword query on any facts at the Internet. Voice seek is regularly interactive, related to several rounds of interaction that lets in a machine to invite for explanation. Voice search is a form of conversation gadget. Voice seek is a speech recognition era that permits customers to search through pronouncing phrases aloud as opposed to typing them right into a search discipline. The proliferation of smart phones and other small, Web-enabled cell devices has spurred interest in voice search. Applications of voice seek include:

- Making search engine queries.
- Clarifying specifics of the request.
- Requesting particular information, inclusive of an inventory quote or sports rating.
- Launching applications and deciding on alternatives.

The unfastened voice seek carrier, but, makes use of another method.  It would possibly appear obvious, but human beings search differently the use of voice than when they kind in a question. Speech popularity and technology technologies provide an ability way to these troubles via augmenting the abilities of an internet browser. The person can talk with the computer and the computer will respond to the consumer inside the shape of voice. The computer will assist the consumer in studying the files as well. The proposed framework is shown in fig 3.

**Figure – 3:** *Proposed Framework*
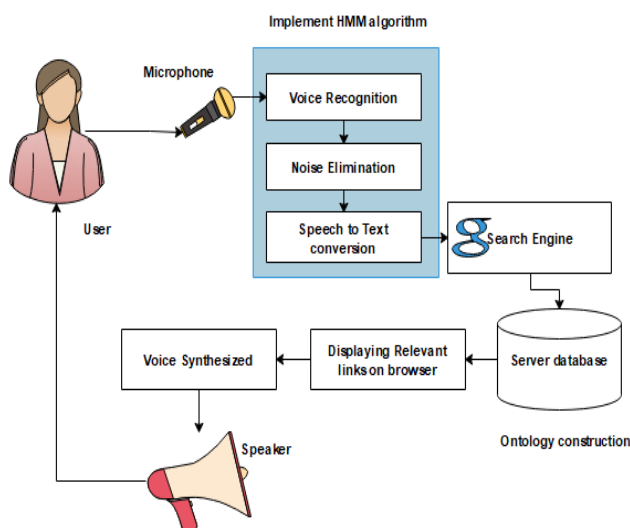
## 4.1 Voice Recognition

In this module the input is given through voice. The voice recognition module compare the given voice based on the pronunciation with the loaded grammar and return the respective action assigned to words such as New Tab, Back, Forward, Print Page and operations like Redirecting the mouse pointer to the URI or search box will help to type words by pronouncing each letters. Alternatively referred to as speech recognition, voice recognition is a computer software program or hardware device with the ability to decode the human voice. Voice recognition is commonly used to operate a device, perform commands, or write without having to use a keyboard, mouse, or press any buttons. Today, this is done on a computer with automatic speech recognition (ASR) software programs. Many ASR programs require the user to "train" the ASR program to recognize their voice so that it can more accurately convert the speech to text.

## 4.2 Speech to Text Conversion

Speech to text conversion is the process of converting spoken words into written texts. This process is also often called speech recognition. Although these terms are almost synonymous, Speech recognition is sometimes used to describe the wider process of extracting meaning from speech, i.e. speech understanding. The term voice recognition should be avoided as it is often associated to the process of identifying a person from their voice, i.e. speaker recognition. All speech-to-text systems rely on at least two models: an acoustic model and a language model. In addition, large vocabulary systems use a pronunciation model. It is important to understand that there is no such thing as a universal speech recognizer. To get the best transcription quality, all of these models can be specialized for a given language, dialect, application domain, type of speech, and communication channel. The complete voice-to-text conversion process is done in three steps. The software first identifies the audio segments containing speech, and then it recognizes the language being spoken if it is not known a priori, and finally it converts the speech segments to text and time-codes.

The basic speech conversion steps as follows:

**Figure – 4:** *Speech Conversion*

$X_t$ : hidden state variables

$y_{ti}$ : $i^{th}$ observed variable @ t
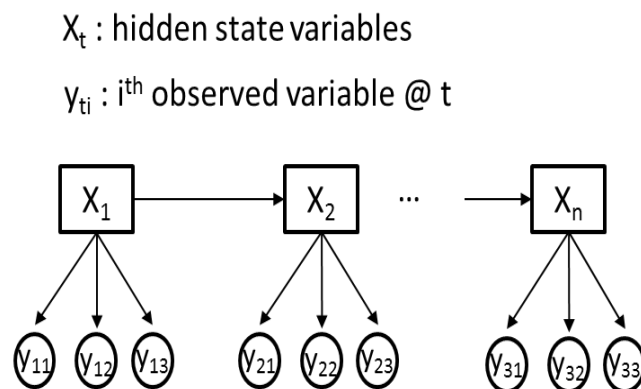


## 4.3 Database Connectivity

In computer science, a database connection is the means by which a database server and its client software communicate with each other. The term is used whether or not the

client and the server are on different machines. The client uses a database connection to send commands to and receive replies from the server. A database is stored as a file or a set of files on magnetic disk or tape, optical disk, or some other secondary storage device. The information in these files may be broken down into records, each of which consists of one or more fields. Fields are the basic units of data storage, and each field typically contains information pertaining to one aspect or attribute of the entity described by the database. Records are also organized into tables that include information about relationships between its various fields. Although database is applied loosely to any collection of information in computer files, a database in the strict sense provides cross-referencing capabilities. Once a connection has been built, it can be opened and closed at will, and properties (such as the command time-out length, or transaction, if one exists) can be set. The connection string consists of a set of key-value pairs, dictated by the data access interface of the data provider.

## 4.4 Display Results

The extraction of text and presenting it to a visually handicapped person has many difficult aspects to it. With innumerous web pages present in the web, there is a varied diversity in the type of the pages. A web page may contain more than one kind of contents like, links, images, advertisements, and animations. These contents may not provide valuable information to a visually impaired person. Further, the document structure of an email page is also different from other page.

## 4.5 Speech Synthesis

The Text To Speech module responds to user like an artificial Intelligent agent by guiding through browsing. This module reads out the content in the web browser by parsing the web document by removing html tags and extracts only text to the user as audio. It also returns information's like date, day, time, weather, etc., on request.

## CONCLUSION

In this paper we proposed an efficient way of getting access to the net browser is presented that's termed as voice surfing wherein visually impaired humans can get entry to the browser the use of speech. As access to net visually incurs obstacles along with visually impaired people can't use keypads, contact screens etc. For giving inputs to pc. User can speech the word and transformed into text robotically, now this browser reduces their effort by using acting this conversion robotically. And the blind people also can use this browser to convert text files in English characters. Thus aggregate of browsing with speech era is an effective way of getting access to webs. This technique can be similarly improvised for a browser that permits visually impaired novices to interact more efficiently with the browser by way of changing their English characters to speech i.e. listening of characters, which can be without problems understood by using them. In addition, all of the text content material gift over the internet for various links can be made accessible by means of the usage of speech era. This technology can also be applied in browser. More paintings may be completed to increase the accuracy, pronunciation and precision of speech generation. The proposed technique has used only English language.

**REFERENCES**

[1] Furui, Sadaoki. "Recent progress in corpus-based spontaneous speech recognition". IEICE Transactions on Information and Systems 88.3 (2005): 366-375.

[2] Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". IEEE Signal processing magazine 29.6 (2012): 82-97.

[3] Real, Esteban, et al. "Large-scale evolution of image classifiers". arXiv preprint arXiv:1703.01041 (2017).

[4] Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams. "Practical bayesian optimization of machine learning algorithms". Advances in neural information processing systems. 2012.

[5] Shahriari, Bobak, et al. "Taking the human out of the loop: A review of bayesian optimization". Proceedings of the IEEE104.1 (2016): 148-175.

[6] Snyder, David, Daniel Garcia-Romero, and Daniel Povey. "Time delay deep neural network-based universal background models for speaker recognition". Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop on. IEEE, 2015.

[7] Dahl, George E., Tara N. Sainath, and Geoffrey E. Hinton. "Improving deep neural networks for LVCSR using rectified linear units and dropout". Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on.IEEE, 2013.

[8] Watanabe, Shinji, and Jonathan Le Roux. "Black box optimization for automatic speech recognition". ICASSP. 2014.

[9] Shinozaki, Takahiro, and Shinji Watanabe. "Structure discovery of deep neural network based on evolutionary algorithms". Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, 2015

[10] Chandrashekaran, Akshay, and Ian Lane. "Automated optimization of decoder hyper-parameters for online lvcsr". Spoken Language Technology Workshop (SLT), 2016 IEEE.IEEE, 2016.